FACTOR ANALYSIS

DR. HEMAL PANDYA
INTRODUCTION

Factor analysis is an INTERDEPENDENCE TECHNIQUE. Thus, there is no distinction between dependent and independent variables.

FA is used to analyze interrelationships among a large number of variables and to explain these variables in terms of their common underlying dimension i.e. Factor.

Thus, analysis can play a unique role in the application of other multivariate techniques.
INTRODUCTION

Factor analysis is a very useful method of reducing data complexity by reducing the number of variables being studied.

“What are the underlying significant drivers of customer’s behavior?”

FA is a good way of resolving this confusion and identify latent or underlying factors from an area of seemingly important variables.

Thus, it is also called DATA REDUCTION technique.
FACTOR ANALYSIS

Factor Analysis is a data reduction technique is used to reduce large set of related variables to a more manageable number, prior to using them in other analysis.

Using Factor Analysis, the researcher can reduce the large number of variables into a few dimensions called factors that summarize the available data.

It aims at grouping the original input variables into factors which underlying the input variables.
INTRODUCTION TO FACTOR ANALYSIS

Factor analysis is a multivariate statistical technique in which there is no distinction between dependent and independent variables.

In factor analysis, all variables under investigation are analyzed together to extract the underlined factors.

Factor analysis is a data reduction method.

It is a very useful method to reduce a large number of variables resulting in data complexity to a few manageable factors.

These factors explain most part of the variations of the original set of data.

A factor is a linear combination of variables.

It is a construct that is not directly observable but that needs to be inferred from the input variables.

The factors are statistically independent.
EXAMPLE

- Data Reduction technique
  - Customer’s ratings of a fast food restaurant.
    - Food taste
    - Food temperature
    - Freshness
    - Waiting time
    - Cleanliness
    - Friendliness of employees
  - Food quality
  - Service quality
KEY TERMS USED IN FACTOR ANALYSIS

Factor:
Linear combinations of original variables.

Factor Scores – It is the composite score estimated for each respondent on the extracted factors.

Factor Loading – The correlation coefficient between the factor score and the variables included in the study is called factor loading. Correlations between original variable and the factors.

Factor Matrix (Component Matrix) – It contains the factor loadings of all the variables on all the extracted factors.

Eigenvalue – The percentage of variance explained by each factor can be computed using eigen value. The eigenvalue of any factor is obtained by taking the sum of squares of the factor loadings of each component. Column sum of squared loadings for a factor; also referred to as a latent root. It represents the amount of variance accounted for by a factor.
KEY TERMS USED IN FACTOR ANALYSIS

Communality:-

Communality - It indicates how much of each variable is accounted for by the underlying factors taken together. In other words, it is a measure of the percentage of variable’s variation that is explained by the factors. Total amount of variance an original variable shares with all other variables included in the analysis.

❖ Exploratory factor analysis:
In EFA data is simply explored and provides information about the numbers of factors required to represent the data.

❖ Confirmatory factor analysis: In confirmatory factor analysis (CFA), researchers can specify the number of factors required in the data and which measured variable is related to which latent variable.

❖ Common factor analysis:-
The extracted factors are based only on the common variance with variable specific and error variance excluded.
**KEY TERMS USED IN FACTOR ANALYSIS**

- **Factor Rotation**
  - Process of manipulation or adjusting the factor axes.

- **Orthogonal**:
  - Mathematical independence (no correlation) of factor axes to each other. (i.e., at right angle, or 90 degrees).

- **Varimax Method**.
  - An orthogonal rotation method that minimizes the number of variables that have high loadings on each factor. This method simplifies the interpretation of the factors.

- **Quartimax Method**.
  - A rotation method that minimizes the number of factors needed to explain each variable. This method simplifies the interpretation of the observed variables.
KEY TERMS USED IN FACTOR ANALYSIS

• **Equamax Method.**
  A rotation method that is a combination of the varimax method, which simplifies the factors, and the quartimax method, which simplifies the variables. The number of variables that load highly on a factor and the number of factors needed to explain a variable are minimized.

• **Direct Oblimin Method.**
  A method for oblique (non-orthogonal) rotation. When delta (Correlation coefficient) equals 0 (the default), solutions are most oblique. As delta becomes more negative, the factors become less oblique. To override the default delta of 0, enter a number less than or equal to 0.8.

• **Promax Rotation**
  An oblique rotation, which allows factors to be correlated. This rotation can be calculated more quickly than a direct oblimin rotation, so it is useful for large datasets.
FACTOR ANALYSIS

General Concepts

*Factor analysis* provides information about reliability, item quality, and construct validity.

General goal is to understand whether and to what extent items from a scale may reflect an underlying hypothetical construct or constructs, known as *factors*.

An analytic method with high sensitivity to identify problematic items and assess the number of factors.
General Concepts

In general, factor analysis methods decompose (or break down) the covariation among items in a measure into meaningful components.

Higher inter-item correlations should reflect greater overlap in what the items measure, and, therefore, higher inter-item correlations reflect higher internal reliability.
FACTOR ANALYSIS

General Concepts

Classical Test Theory (CTT)

\[
\begin{align*}
\text{Observed Score} & = \text{True Score} + \text{Error} \\
X_o & = X_t + X_e
\end{align*}
\]
FACTOR ANALYSIS

General Concepts

Factor model concept is analogous to CTT

True score $X_t$  
Factor  
Observed score $X_o$  
Measured variable  
Error $X_e$  
Error (measurement residual)
General Concepts

In practice, a factor cannot be estimated with one item. Should only be estimated with three or more items. Items with higher correlation with factor contribute more to the measure.
FACTOR ANALYSIS

General Concepts

Items are referred to as *indicators*
Regression slopes between factor and indicators are referred to as *loadings*
FACTOR ANALYSIS

General Concepts

Patterns of high inter-item correlations among subsets of items suggest more than one factor because the items tend to "cluster" together.

Any number of factors might underlie a set of items, up to the total number of items (which would imply no common factor).

Example: set of six items might assess extroversion and openness.
FACTOR ANALYSIS

General Concepts

We never know the meaning of the factors, however; we can only use theory to decide what they mean and then test their validity.

The factors may be related or not related—correlated or orthogonal (uncorrelated).

If those who are extroverted tend to be a little more open, then the factors are correlated (contrary to what is suggested by the table).
SAMPLE SIZE

In small samples, the correlation coefficients among the variables are less reliable, tending to vary from sample to sample.

So, Nunnally (1978) recommends a 10 to 1 ratio; that is, ten cases for each item to be factor analyzed.

Others suggest that five cases for each are adequate in most cases.

A common rule is to suggest that a researcher has at least 10-15 participants per variable.

Fiedel (2005) says that in general over 300 respondents for sampling analysis is probably adequate.

There is universal agreement that factor analysis is inappropriate when sample size is below 50.
Principal Component Analysis (PCA)

Exploratory Factor Analysis (EFA) and

Confirmatory Factor Analysis (CFA).

PCA and EFA are used in the early stages of research to gather information about (explore) the interrelationships among a set of variables. But, another technique confirmatory analysis is a set of technique used later in the research process to test (confirm) specific hypotheses or theories concerning the structure underlying a set of variables.
Principal Component Analysis (PCA) is a data reduction technique that transforms a large number of correlated variables into a much smaller set of uncorrelated variables called principal components.

Principal Components are the linear combinations of the observed variables. The weights used to form the linear composites are chosen to maximize the variance each principal component accounts for, while keeping the components uncorrelated.
If we take an example, we have 6 observed variables (S₁ to S₆) and PC₁ and PC₂ are any two derived variables i.e., principal components, then each component is a linear combination of observed variables. i.e., \( PC₁ = \alpha₁ S₁ + \alpha₂ S₂ + \cdots + \alpha₆ S₆ \) and \( PC₂ = \beta₁ S₁ + \beta₂ S₂ + \cdots + \beta₆ S₆ \)

Diagrammatic representation is given below:

Here, we have to take each variable as standardized variables. From this we understand, the first principal component is the weighted combination of the 6 observed variables that accounts for the most variance in the original set of variables. The second principal component is the linear combination that accounts for the most variance in the original variables, under the constraint that it’s orthogonal (uncorrelated) to the first principal component. Each subsequent component maximizes the variance accounted for, while at the same time remaining uncorrelated with all previous components.
STEPS FOR HANDLING DATA IN PRINCIPAL COMPONENT ANALYSIS:
Check the inter-correlations among the items using KMO and Bartlett’s test of Sphericity

Factor Extraction:
It involves determining the smaller number of factors that can be used to best represent the interrelations among the set of variables. Many approaches are available to extract some underlying factors using observed variables. Some methods are

Kaiser’s criterion: This approach is based on eigenvalue. Each component is associated with an eigenvalue of the correlation matrix. The eigenvalue of a factor represents the amount of total variance explained by the factor. Using this rule, only factors with an eigenvalue of 1 or more are retained for further investigation because Components with eigen values less than 1 explain less variance than contained in a single variable.
STEPS FOR HANDLING DATA IN PRINCIPAL COMPONENT ANALYSIS:

Scree test: Another approach that can be used is Cattell scree test (Cattell, 1966). In this test, the eigen values are plotted against their component numbers. Such plots will typically demonstrate a bend or elbow, and the components above this sharp break are retained.

Parallel analysis: Parallel analysis (Horn 1965) involves comparing the size of the eigen values with those obtained from a randomly generated data set of the same size. Only those eigen values that exceed the average corresponding eigen values from the random data set are retained.
STEPS FOR HANDLING DATA IN PRINCIPAL COMPONENT ANALYSIS:

Factor rotation: If the number of factors determined are more than 1 then factor rotation needs. Rotations are a set of mathematical techniques for transforming the component loading matrix into one that’s more interpretable. It presents the pattern of loadings in a manner that is easier to interpret. There are two main approaches to rotation, resulting in either orthogonal (uncorrelated) or oblique (correlated). In PCA method, we use most popular orthogonal method “varimax” rotation, which attempts to purify the columns of the loading matrix, so that each component is defined by a limited set of variables. It minimizes the number of variables that have high loading on each factor.
CONDITIONS FOR A FA EXERCISE

• **FA exercise requires metric data. i.e.** the data should be either interval or ratio scale in nature.

• The variables for FA are identified through exploratory research. (By Literature Review or informal interviews, focused group discussion etc.) (Generally in survey research 5/7 point Likert scale may be used.)

• As the responses to different statements are obtained through different scales, all the responses need to be standardized.

• The size of sample respondents should be at least **four to five times more** than the number of variables (no. of statements)
CONDITIONS FOR A FA EXERCISE

• The basic principle behind FA is that the initial set of variables should be highly correlated.

• If the correlation coefficients between all the variables are small, FA may not be appropriate technique.

• For application of FA the value of KMO (KAISER-MEYER OLKIN)- should be greater than 0.5.

• The KMO statistic compares the magnitude of observed correlation coefficients with the magnitude of partial correlation coefficients. (RANGE OF KMO STATISTIC ➔ 0 to 1)

• Bartlett statistic for sphericity: Bartlett's test of sphericity tests the hypothesis that your correlation matrix is an identity matrix, which would indicate that your variables are unrelated and therefore unsuitable for structure detection.
CONDITIONS FOR A FACTOR ANALYSIS EXERCISE

The following conditions must be ensured before executing the technique:

Factor analysis exercise requires metric data. This means the data should be either interval or ratio scale in nature.

The variables for factor analysis are identified through exploratory research which may be conducted by reviewing the literature on the subject, researches carried out already in this area, by informal interviews of knowledgeable persons, qualitative analysis like focus group discussions held with a small sample of the respondent population, analysis of case studies and judgment of the researcher.

As the responses to different statements are obtained through different scales, all the responses need to be standardized. The standardization helps in comparison of different responses from such scales.
CONDITIONS FOR A FACTOR ANALYSIS EXERCISE

The size of the sample respondents should be at least four to five times more than the number of variables (number of statements).

The basic principle behind the application of factor analysis is that the initial set of variables should be highly correlated. If the correlation coefficients between all the variables are small, factor analysis may not be an appropriate technique.

The significance of correlation matrix is tested using Bartlett’s test of sphericity. The hypothesis to be tested is

\[ H_0 : \text{Correlation matrix is insignificant, i.e., correlation matrix is an identity matrix where diagonal elements are one and off diagonal elements are zero.} \]

\[ H_1 : \text{Correlation matrix is significant.} \]
CONDITIONS FOR A FACTOR ANALYSIS EXERCISE

The test converts it into a chi-square statistics with degrees of freedom equal to $[(k(k-1))/2]$, where $k$ is the number of variables on which factor analysis is applied. The significance of the correlation matrix ensures that a factor analysis exercise could be carried out.

The value of Kaiser-Meyer-Olkin (KMO) statistics which takes a value between 0 and 1 should be greater than 0.5 for the application of factor analysis.

The KMO statistics compares the magnitude of observed correlation coefficients with the magnitudes of partial correlation coefficients.

A small value of KMO shows that correlation between variables cannot be explained by other variables.
STEPS IN A FACTOR ANALYSIS EXERCISE

There are basically two steps that are required in a factor analysis exercise.

Extraction of factors:

The first and the foremost step is to decide on how many factors are to be extracted from the given set of data. The principal component method is discussed very briefly here.

As we know that factors are linear combinations of the variables which are supposed to be highly correlated, the mathematical form of the same could be written as

\[ F_i = W_{i1}X_1^* + W_{i2}X_2^* + W_{i3}X_3^* + \ldots + W_{ik}X_k^* \]

Where,
- \( X_i \) = \( i^{th} \) standardized variable
- \( F_i \) = Estimate of \( i^{th} \) factor
- \( W_i \) = Weight or factor score coefficient for \( i^{th} \) standardized variable.
- \( k \) = Number of variables
The principal component methodology involves searching for those values of $W_i$ so that the first factor explains the largest portion of total variance. This is called the first principal factor.

This explained variance is then subtracted from the original input matrix so as to yield a residual matrix.

A second principal factor is extracted from the residual matrix in a way such that the second factor takes care of most of the residual variance.

One point that has to be kept in mind is that the second principal factor has to be statistically independent of the first principal factor. The same principle is then repeated until there is little variance to be explained.
STEPS IN A FACTOR ANALYSIS EXERCISE

To decide on the number of factors to be extracted, Kaiser Guttman methodology is used which states that the number of factors to be extracted should be equal to the number of factors having an eigenvalue of at least 1.

Rotation of factors:

The second step in the factor analysis exercise is the rotation of initial factor solutions. This is because the initial factors are very difficult to interpret. Therefore, the initial solution is rotated so as to yield a solution that can be interpreted easily.

The varimax rotation method is used.
The varimax rotation method maximizes the variance of the loadings within each factor.

The variance of the factor is largest when its smallest loading tends towards zero and its largest loading tends towards unity.

The basic idea of rotation is to get some factors that have a few variables that correlate high with that factor and some that correlate poorly with that factor.

Similarly, there are other factors that correlate high with those variables with which the other factors do not have significant correlation.

Therefore, the rotation is carried out in such way so that the factor loadings as in the first step are close to unity or zero.
**STEPS IN A FACTOR ANALYSIS EXERCISE**

To interpret the results, a cut-off point on the factor loading is selected.

There is no hard and fast rule to decide on the cut-off point. However, generally it is taken to be greater than 0.5.

All those variables attached to a factor, once the cut-off point is decided, are used for naming the factors. This is a very subjective procedure and different researchers may name same factors differently.

A variable which appear in one factor should not appear in any other factor. This means that a variable should have a high loading only on one factor and a low loading on other factors.
If that is not the case, it implies that the question has not been understood properly by the respondent or it may not have been phrased clearly.

Another possible cause could be that the respondent may have more than one opinion about a given item (statement).

The total variance explained by Principal component method and Varimax rotation is same. However, the variance explained by each factor could be different.

The communalities of each variable remains unchanged by both the methods.
STEPS IN FA EXERCISE

• There are two stages in FA.

STAGE 1: FACTOR EXTRACTION PROCESS

• Identify how many factors will be extracted from the data.

• This could be accomplished by various methods like the centroid method, the principal component method and maximum likelihood method.

• There is a rule of thumb based on the computation of EIGEN VALUES, to determine how many factors to extract.
STEPS IN FA EXERCISE

• The Eigen value for a given factor measures the variance in all the variables which is accounted for by that factor.

• Eigen values measure the amount of variation in the total sample accounted for by each factor.

• If a factor has a low Eigen value, then it is contributing little to the explanation of variances in the variables and may be ignored as redundant with more important factors.

• In short, we can have as many factors as there are original variables. But the objective is to reduce the variables to a fewer number of factors, RETAIN those WITH EIGEN VALUE of 1 or more.

• (Before extraction it is assumed that each of the original variable has an Eigen value = 1)
STEPS IN FA EXERCISE

STAGE : 2 ROTATION OF FACTORS / PRINCIPAL COMPONENT

This is actually optional, but highly recommended.

After the number of extracted factors is decided in stage 1, the next task of the researcher is to interpret and name the factor. The factor matrix is used for this purpose.

The original factor matrix is un-rotated and is a part of output from stage 1.

The rotated factor matrix comes about in stage 2.

Most of the computer software would give options for orthogonal rotation, Varimax rotation and Oblique rotation.

Generally, the Varimax rotation is used as this results in independent factors.
Stage 1

➢ Research Problem
• Is the analysis exploratory or confirmatory?
➢ Select objectives:
  • Data summarization and Data reduction

Confirmatory
Structural equation modeling

Exploratory

To stage 2
Stage 2

Research Design
- What variables are included?
- How are the variables measure?
- What is the desired sample size?

Stage 3

Assumptions
Statistical consideration of NORMALITY, LINEARITY and HOMOSCEDASTICITY.
HOMOGENETY OF SAMPLE

From stage 1

To stage 4
Stage 4

Selecting a factor method:
Is the total variance or only common variance analyzed?

Total Variance
Extract factors with component analysis

Common Variance
Extract factors with common factor analysis

Specifying the factor matrix
Determine the number of factor to be retained

To stage 5
Stage 5

Selecting a Rotational method:
Should the factors be correlated(oblique) or uncorrelated(orthogonal)?

Orthogonal Methods:
- VARIMAX
- EQUIMAX
- QUARTIMAX

Oblique Methods:
- Oblimin
- Promax
- Orthoblique

Interpreting the Rotated Factor Matrix:
- Can significant loadings be found?
- Do you want to change number of factors?
- Are communalities sufficient?

Yes (Go to stage 6)

No (Go to stage 4)
Stage 6

Validation of the factor matrix
- Split/multiple samples
- Separate analysis of subgroups
- Identify influential cases

Stage 7

Selection of surrogate variables
Computation of factor scores
Creation of summated scales
EXPLORATORY FACTOR ANALYSIS

In multivariate statistics, exploratory factor analysis (EFA) is a statistical method used to uncover the underlying structure of a relatively large set of variables. EFA is a technique within factor analysis whose overarching goal is to identify the underlying relationships between measured variables.

The partitioning of variance differentiates a principal components analysis from what we call common factor analysis. Both methods try to reduce the dimensionality of the dataset down to fewer unobserved variables, but whereas PCA assumes that there common variances takes up all of total variance, common factor analysis assumes that total variance can be partitioned into common and unique variance. It is usually more reasonable to assume that you have not measured your set of items perfectly.
EXPLORATORY FACTOR ANALYSIS

Exploratory Factor Analysis

Two major types of factor analysis

*Exploratory factor analysis* (EFA)

*Confirmatory factor analysis* (CFA)

Major difference is that EFA seeks to discover the number of factors and does not specify which items load on which factors
Exploratory Factor Analysis

In EFA, loadings are obtained for all items related to all factors
EXPLORATORY FACTOR ANALYSIS

Exploratory Factor Analysis

The researcher may discover there is one factor underlying the items or many factors. Items may be eliminated by the researcher if they do not load highly. Researchers choose items that load highly on one factor and low on other factors to achieve simple structure. Composite scale scores often created based on the factor analysis to be used in further research.
EXPLORATORY FACTOR ANALYSIS

Exploratory Factor Analysis

EFA is available in most general statistical software, such as SPSS, R, SAS

Involves several steps and decision points
  Deciding on the number of factors
  Extraction
  Rotation
Exploratory Factor Analysis

An initial analysis called *principal components analysis* (PCA) is first conducted to help determine the number of factors that underlie the set of items.

PCA is the default EFA method in most software and the first stage in other exploratory factor analysis methods to select the number of factors.

PCA is not considered a “true factor analysis method,” because measurement error is not estimated (Snook & Gorsuch, 1989).
Exploratory Factor Analysis

PCA gives eigenvalues for the number of components (factors) equal to the number of items.
If 12 items, there will be 12 eigenvalues.
Each component is a potential “cluster” of highly inter-correlated items.
Eigenvalues represent the amount of variance accounted for by each component, but they are not in a standardized metric.
Larger eigenvalues indicate a more important (and more likely real) components or factor, with some merely reflecting unimportant factors or random variation.
Exploratory Factor Analysis

The values sum to the number of items, so if 12 items, then there will be 12 eigenvalues that sum to 12.

The proportion or percentage of (co)variance accounted for by each factor can be calculated by dividing by the number of items.
Exploratory Factor Analysis

There are several possible rules which may be used for choosing the number of factors based on eigenvalues. The usual rule of greater than 1.0 (the Kaiser-Guttman rule) does not seem to work the best (Preacher & MacCallaum, 2003).

Most use the scree plot and a *subjective scree test* by identifying the biggest drop in eigenvalues. The scree test or a more objective version (Cattell–Nelson–Gorsuch test) seems to work well for identifying the correct number of factors (Cattell & Vogelmann, 1977).
Exploratory Factor Analysis

Exploratory Factor Analysis

Next steps in an EFA after deciding on the number of factors is to choose a *method of extraction*. The extraction method is the statistical algorithm used to estimate loadings. There are several to choose from, of which *principal factors* (principal axis factoring) or *maximum likelihood* seem to perform the best (Fabrigar et al., 1999).
### Correspondence Index for Exploratory Factor Analysis

<table>
<thead>
<tr>
<th>Indicators</th>
<th>Cut-off Value</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kaiser–Meyer–Olkin (KMO)</td>
<td>Recommended value of 0.6 or above</td>
<td>Hair et al. (2010)</td>
</tr>
<tr>
<td>Meritorious: ≥0.80,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Middling: ≥0.70,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mediocre: ≥0.60,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miserable: ≥0.50,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unacceptable: &lt;0.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bartlett’s Test of Sphericity</td>
<td>Significant at $\alpha &lt; .05$</td>
<td>Hair et al. (2010)</td>
</tr>
<tr>
<td>Anti-Image Correlation:</td>
<td>$&gt; 0.5$</td>
<td>Coakes &amp; Steed, (2003); Hair et al., (2010)</td>
</tr>
<tr>
<td>individual measure of sampling adequacy (MSA)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Communalities (variables are well defined by the solution—low values require removal)</td>
<td>$&gt; 0.3$</td>
<td>Tabachnick &amp; Fidell (2007); Gaskin (2012); Hair et al. (2010)</td>
</tr>
<tr>
<td></td>
<td>$&gt; 0.4$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$&gt; 0.5$</td>
<td></td>
</tr>
<tr>
<td>Factor loadings Significant Factor Loading based on Sample Size</td>
<td>Above sufficient factor loading to retain the item while below sufficient factor loading to eliminate the item.</td>
<td>Hair et al. (2010)</td>
</tr>
</tbody>
</table>
Exploratory Factor Analysis

And factor rotation
Factor rotation is a mathematical scaling process for the loadings that also specifies whether the factors are correlated (oblique) or uncorrelated (orthogonal)
Usually no harm in allowing factors to correlate
If the factor correlation is zero, then the same as orthogonal
Orthogonal rotation makes a strong assumption that the factors are uncorrelated, which probably is not likely in most applications
USES OF FACTOR ANALYSIS

Scale construction: Factor analysis could be used to develop concise multiple item scales for measuring various constructs.

Establish antecedents: This method reduces multiple input variables into grouped factors. Thus, the independent variables can be grouped into broad factors.

Psychographic profiling: Different independent variables are grouped to measure independent factors. These are then used for identifying personality types.

Segmentation analysis: Factor analysis could also be used for segmentation. For example, there could be different sets of two-wheelers-customers owning two-wheelers because of different importance they give to factors like prestige, economy consideration and functional features.
USES OF FACTOR ANALYSIS

Marketing studies: The technique has extensive use in the field of marketing and can be successfully used for new product development; product acceptance research, developing of advertising copy, pricing studies and for branding studies.

For example we can use it to:

- identify the attributes of brands that influence consumers’ choice;
- get an insight into the media habits of various consumers;
- identify the characteristics of price-sensitive customers.
USES OF FA

 SCALE CONSTRUCTION:

- Example: - 15-item scale to measure job satisfaction.
- At first step - Generate large number of statements, numbering say 100 or so as a part of exploratory research.
- Assume that we get 3 factors out of it.
- I want to construct 15- item scale to measure job satisfaction.
- Separate 5 items from each factors having highest factor loading.
USES OF FA

- Establish antecedents: (Data summarization)
  - This method reduces multiple input variables into grouped factors.

  • For example, All the variables that measure safety clauses in mutual fund could be reduced to a factor called SAFETY CLAUSE.
USES OF FA

Psychographic profiling:

- **Psychographics** can be defined as a quantitative methodology used to describe consumers on psychological attributes.

- When a relatively complete profile of a person or group's psychographic make-up is constructed, this is called a "psychographic profile".

- Some categories of psychographic factors used in market segmentation include:
  - activity, interest, opinion (AIOs)
  - attitudes
  - values
  - behavior
USES OF FA

- **Segmentation analysis:**
  - Factor analysis could also be used for segmentation.
  - Example: There could be different sets of two-wheelers customers owing two-wheelers because of different importance they give to factors like:
    - Prestige,
    - Economy consideration,
    - Functional features,
    - Traffic and time etc.
USES OF FA

- Marketing studies:
  - This technique has extensive use in the field of marketing and can be successfully used for new product development; product acceptance research, developing advertising copy, pricing studies, branding studies and so on….
  - For example it can be used to…
    - identify the attributes of brands that influence consumer’s choice.
    - identify the characteristics of price sensitive customers etc..
APPLICATIONS OF FACTOR ANALYSIS IN OTHER MULTIVARIATE TECHNIQUES

1. Multiple regression – Factor scores can be used in place of independent variables in a multiple regression estimation. This way we can overcome the problem of multicollinearity.

2. Simplifying the discrimination solution – A number of independent variables in a discriminant model can be replaced by a set of manageable factors before estimation.

3. Simplifying the cluster analysis solution - To make the data manageable, the variables selected for clustering can be reduced to a more manageable number using a factor analysis and the obtained factor scores can then be used to cluster the objects/cases under study.

4. Perceptual mapping in multidimensional scaling - Factor analysis that results in factors can be used as dimensions with the factor scores as the coordinates to develop attribute-based perceptual maps where one is able to comprehend the placement of brands or products according to the identified factors under study.
In contrast to PCA, factors F1, F2 and F3 are assumed to underlie or “cause” the observed variables, rather than being linear combinations of them. The errors $e_1$ to $e_8$ represent the variance in the observed variables unexplained by the factors. The circles indicate that the factors and errors are not directly observable but are inferred from the correlations among the variables. Here curved arrow between the factors indicates that they are correlated. Correlated factors are common but required, in the EFA model.

Each factor is assumed to explain the variance shared among two or more observed variables, so technically, they are called common factors.

The model can be represented as

$$S_i = \alpha_{i1}F_1 + \alpha_{i2}F_2 + \alpha_{i3}F_3 + U_i$$

Where $S_i$ is the $i$th observed variable ($i=1, 2, \ldots, 8$), $F_j$ are the common factors ($j=1, 2, 3$) and $U_i$ is the portion of variable $S_i$ unique to that variable (not explained by the common factors). The $\alpha_{ij}$ can be thought of as the degree to which each factor contributes to the composition of an observed variable.

The same procedure which is used for PCA can be processed for EFA except the method (principal component) for extraction. In EFA, Principal axis factoring method can be used.
SPSS COMMANDS FOR FACTOR ANALYSIS

Exploratory Factor Analysis

Analyze>Dimension
Reduction>Factor>Descriptive>Extraction>Based on Eigen Value>Rotation>Scores>Options>OK

Confirmatory Factor Analysis

Analyze>Dimension
Reduction>Factor>Descriptive>Extraction>Fixed Number of Factors>Specify Number of Factors to be Extracted>Rotation>Scores>Options>OK
CASE STUDY

20 two-wheeler users were surveyed about their perception and image attributes of vehicles they owned.

Ten statements were as follows:

1. I use a two-wheeler because it is affordable.
2. It gives me sense of freedom to own a two-wheeler.
3. Low maintenance cost makes a two-wheeler very economical in the long run.
4. A two-wheeler is essentially a man’s vehicle.
5. I feel very powerful when I am on my two-wheeler.
6. Some of my friends who don’t have their own vehicle are jealous of me.
7. I feel good whenever I see the ad for my two-wheeler on TV, in a magazine or on a hording.
8. My vehicle gives me a comfortable ride.
9. I think two-wheelers are safe way to travel.
10. Three people should be legally allowed to travel on a two-wheeler.
RESULTS

Descriptive Statistics

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>S. D.</th>
<th>N</th>
<th>C.V.</th>
</tr>
</thead>
<tbody>
<tr>
<td>VAR00001</td>
<td>2.35</td>
<td>1.3089</td>
<td>20</td>
<td>55.7</td>
</tr>
<tr>
<td>VAR00002</td>
<td>3.25</td>
<td>1.4824</td>
<td>20</td>
<td>45.6</td>
</tr>
<tr>
<td>VAR00003</td>
<td>2.25</td>
<td>1.118</td>
<td>20</td>
<td>49.7</td>
</tr>
<tr>
<td>VAR00004</td>
<td>3.1</td>
<td>1.8035</td>
<td>20</td>
<td>58.2</td>
</tr>
<tr>
<td>VAR00005</td>
<td>2.8</td>
<td>1.5079</td>
<td>20</td>
<td>53.9</td>
</tr>
<tr>
<td>VAR00006</td>
<td>3.05</td>
<td>1.6051</td>
<td>20</td>
<td>52.6</td>
</tr>
<tr>
<td>VAR00007</td>
<td>2.7</td>
<td>1.4546</td>
<td>20</td>
<td>53.9</td>
</tr>
<tr>
<td>VAR00008</td>
<td>3.05</td>
<td>1.905</td>
<td>20</td>
<td>62.5</td>
</tr>
<tr>
<td>VAR00009</td>
<td>3.2</td>
<td>1.5079</td>
<td>20</td>
<td>47.1</td>
</tr>
<tr>
<td>VAR00010</td>
<td>2.8</td>
<td>1.4726</td>
<td>20</td>
<td>52.6</td>
</tr>
</tbody>
</table>

Lowest C.V. indicates that particular variable is most consistent. Variable 2 (sense of freedom) is having lowest c.v., thus it is the most consistent variable.
RESULTS

<table>
<thead>
<tr>
<th>KMO and Bartlett's Test</th>
<th>Kaiser-Meyer-Olkin Measure of Sampling Adequacy</th>
<th>.618</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bartlett's Test of Sphericity</td>
<td>Approx. Chi-Square</td>
<td>164.098</td>
</tr>
<tr>
<td></td>
<td>df</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>Sig.</td>
<td>.000</td>
</tr>
</tbody>
</table>

• The **Kaiser-Meyer- Olkin Measure of Sampling Adequacy** is a statistic that indicates the proportion of variance in your variables that might be caused by underlying factors.
• High values (close to 1.0) generally indicate that a factor analysis may be useful with your data.
• If the value is less than 0.50, the results of the factor analysis probably won't be very useful.
• **Bartlett's test of sphericity** tests the hypothesis that your correlation matrix is an identity matrix, which would indicate that your variables are unrelated and therefore unsuitable for structure detection.
• Small values (less than 0.05) of the significance level indicate that a factor analysis may be useful with your data.
Kaiser-Meyer-Olkin (KMO) Test is a measure of how suited your data is for Factor Analysis. The test measures sampling adequacy for each variable in the model and for the complete model. The statistic is a measure of the proportion of variance among variables that might be common variance.
The Kaiser-Mayer-Olkin measure of sample adequacy is

\[ KMO_j = \frac{\sum_{i \neq j} r_{ij}^2}{\sum_{i \neq j} r_{ij}^2 + \sum_{i \neq j} a_{ij}^2} \]

\[ KMO = \frac{\sum_{i \neq j} \sum r_{ij}^2}{\sum_{i \neq j} r_{ij}^2 + \sum_{i \neq j} a_{ij}^2} \]

Where \( a_{ij}^2 \) is the anti-image correlation coefficient (negatives of the partial correlation coefficients) and \( r_{ij} \) is the correlation between \( i \) and \( j \)

Let \( S^2 = \text{diag}(R^{-1})^{-1} \) and \( Q = SR^{-1}S \). Then \( Q \) is said to be the anti-image intercorrelation matrix. Let \( \sum r^2 = \sum R^2 \) and \( \sum q^2 = \sum Q^2 \) for all off diagonal elements of \( R \) and \( Q \), then \( SMA = \sum r^2 / (\sum r^2 + \sum q^2) \). Although originally MSA was \( 1 - \frac{\sum q^2}{\sum r^2} \) (Kaiser, 1970), this was modified in Kaiser and Rice, (1974) to be \( SMA = \frac{\sum r^2}{(\sum r^2 + \sum q^2)} \). This is the formula used by Dziuban and Shirkey (1974).
A RULE OF THUMB FOR INTERPRETING THE STATISTIC

KMO values between 0.8 and 1 indicate the sampling is adequate. KMO values less than 0.6 indicate the sampling is not adequate and that remedial action should be taken. Some authors put this value at 0.5, so use your own judgment for values between 0.5 and 0.6.

The KMO measures the sampling adequacy (which determines if the responses given with the sample are adequate or not) which should be close than 0.5 for a satisfactory factor analysis to proceed.

Kaiser (1974) recommend 0.5 (value for KMO) as minimum (barely accepted), values between 0.7-0.8 acceptable, and values above 0.9 are superb.
Bartlett (1951) introduced the test of sphericity, which tests whether a matrix is significantly different from an identity matrix. This statistical test for the presence of correlations among variables, provides the statistical probability that the correlation matrix has significant correlations among at least some of variables.

Small values (less than 0.05) of the significance level indicate that a factor analysis may be useful with our data.

This tests the null hypothesis that the correlation matrix is an identity matrix.

If we rejected the hypothesis then we can say Bartlett’s test of sphericity suggests that there is sufficient significant correlation in the data for factor analysis.

Statistical test for overall significance of all correlations within a correlation matrix.
TERMINOLOGY IN FACTOR (COMPONENT) ANALYSIS

Factor or Principal Component

Factor(Component) Loadings (correlation coefficients between the variables and the factors)

Variance (Common and Unique)

Common variance is the amount of variance that is shared among a set of items. Items that are highly correlated will share a lot of variance.
Unique variance is any portion of variance that’s not common. There are two types:

- **Specific variance**: is variance that is specific to a particular item
- **Error variance**: comes from errors of measurement and basically anything unexplained by common or specific variance

If the total variance is 1, then the communality is $h^2$ and the unique variance is $1-h^2$.

**Eigen Values**: Eigen values represent the total amount of variance that can be explained by a given principal component. Eigen values are calculated by adding the squares of factor loading of all the variables in the factor. Generally factors with eigen values more than 1.0 are considered stable.
Total Variance explained: The total variance explained is the percentage of total variance of the variables explained. This is calculating by adding all the communality values of each variable and dividing it by the number of variables. It will be 1 for PCA.

Factor Variance explained: The factor variance explained is the percentage of total variance of the variables explained by the factors. This is calculating by adding the squared factor loadings of all the variables (eigenvalue) and dividing it by the number of variables.
## RESULTS

![Total Variance Explained Table](image)

**Extraction Method:** Principal Component Analysis.
RESULTS

The first section of the table shows the Initial Eigen values.
The Total column gives the Eigen value, or amount of variance in the original variables accounted for by each component.

The % of Variance column gives the ratio, expressed as a percentage, of the variance accounted for by each component to the total variance in all of the variables.

The Cumulative % column gives the percentage of variance accounted for by the first \( n \) components.

For example, the cumulative percentage for the second component is the sum of the percentage of variance for the first and second components.
RESULTS

The second section of the table shows the extracted components. They explain nearly 80% of the variability in the original ten variables, so you can considerably reduce the complexity of the data set by using these components, with only a 20% loss of information.
# RESULTS

<table>
<thead>
<tr>
<th>Component Matrix(^a)</th>
<th>Component</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>VAR00001</td>
<td></td>
<td>.176</td>
<td>.670</td>
<td>.493</td>
</tr>
<tr>
<td>VAR00002</td>
<td></td>
<td>-.136</td>
<td>-.608</td>
<td>.254</td>
</tr>
<tr>
<td>VAR00003</td>
<td></td>
<td>-.107</td>
<td>.820</td>
<td>.218</td>
</tr>
<tr>
<td>VAR00004</td>
<td></td>
<td>.966</td>
<td>-.036</td>
<td>-.097</td>
</tr>
<tr>
<td>VAR00005</td>
<td></td>
<td>.951</td>
<td>.166</td>
<td>-.136</td>
</tr>
<tr>
<td>VAR00006</td>
<td></td>
<td>.952</td>
<td>-.084</td>
<td>-.025</td>
</tr>
<tr>
<td>VAR00007</td>
<td></td>
<td>.971</td>
<td>.096</td>
<td>-.046</td>
</tr>
<tr>
<td>VAR00008</td>
<td></td>
<td>-.322</td>
<td>.775</td>
<td>-.308</td>
</tr>
<tr>
<td>VAR00009</td>
<td></td>
<td>-.069</td>
<td>.735</td>
<td>-.482</td>
</tr>
<tr>
<td>VAR00010</td>
<td></td>
<td>.161</td>
<td>.319</td>
<td>.814</td>
</tr>
</tbody>
</table>

Extraction Method: Principal Component Analysis.

a. 3 components extracted.

<table>
<thead>
<tr>
<th>Rotated Component Matrix(^a)</th>
<th>Component</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>VAR00001</td>
<td></td>
<td>.126</td>
<td>.313</td>
<td>.780</td>
</tr>
<tr>
<td>VAR00002</td>
<td></td>
<td>-.181</td>
<td>-.639</td>
<td>-.107</td>
</tr>
<tr>
<td>VAR00003</td>
<td></td>
<td>-.116</td>
<td>.604</td>
<td>.594</td>
</tr>
<tr>
<td>VAR00004</td>
<td></td>
<td>.970</td>
<td>-.064</td>
<td>-.006</td>
</tr>
<tr>
<td>VAR00005</td>
<td></td>
<td>.964</td>
<td>.131</td>
<td>.063</td>
</tr>
<tr>
<td>VAR00006</td>
<td></td>
<td>.945</td>
<td>-.140</td>
<td>.030</td>
</tr>
<tr>
<td>VAR00007</td>
<td></td>
<td>.971</td>
<td>.024</td>
<td>.106</td>
</tr>
<tr>
<td>VAR00008</td>
<td></td>
<td>-.262</td>
<td>.848</td>
<td>.101</td>
</tr>
<tr>
<td>VAR00009</td>
<td></td>
<td>.010</td>
<td>.881</td>
<td>-.044</td>
</tr>
<tr>
<td>VAR00010</td>
<td></td>
<td>.063</td>
<td>-.149</td>
<td>.874</td>
</tr>
</tbody>
</table>

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 5 iterations.
RESULTS

Communalities, denoted by $h^2$, measure the percentage of variance in each variable explained by the factors extracted. It ranges from 0 to 1.

Examine the communality values to assess how well each variable is explained by the factors.

The closer the communality is to 1, the better the variable is explained by the factors.

You can decide to add a factor if the factor contributes significantly to the fit of certain variables.

<table>
<thead>
<tr>
<th>Communalities</th>
<th>Initial</th>
<th>Extraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>VAR00001</td>
<td>1.000</td>
<td>.722</td>
</tr>
<tr>
<td>VAR00002</td>
<td>1.000</td>
<td>.452</td>
</tr>
<tr>
<td>VAR00003</td>
<td>1.000</td>
<td>.731</td>
</tr>
<tr>
<td>VAR00004</td>
<td>1.000</td>
<td>.945</td>
</tr>
<tr>
<td>VAR00005</td>
<td>1.000</td>
<td>.950</td>
</tr>
<tr>
<td>VAR00006</td>
<td>1.000</td>
<td>.914</td>
</tr>
<tr>
<td>VAR00007</td>
<td>1.000</td>
<td>.955</td>
</tr>
<tr>
<td>VAR00008</td>
<td>1.000</td>
<td>.799</td>
</tr>
<tr>
<td>VAR00009</td>
<td>1.000</td>
<td>.777</td>
</tr>
<tr>
<td>VAR00010</td>
<td>1.000</td>
<td>.789</td>
</tr>
</tbody>
</table>

Extraction Method: Principal
## NAME OF THE FACTORS

<table>
<thead>
<tr>
<th>FACTORS</th>
<th>NAME</th>
</tr>
</thead>
<tbody>
<tr>
<td>FACTOR 1</td>
<td>➔4,5,6,7</td>
</tr>
<tr>
<td></td>
<td>‘MALE EGO’ or ‘PRIDE OF OWERNERSHIP’</td>
</tr>
<tr>
<td>FACTOR 2</td>
<td>➔3,8,9</td>
</tr>
<tr>
<td></td>
<td>‘LOW MAINTAINANCE’ or ‘COMFORT’ or ‘SAFETY’</td>
</tr>
<tr>
<td>FACTOR 3</td>
<td>➔1,10</td>
</tr>
<tr>
<td></td>
<td>‘AFFORDABILITY’</td>
</tr>
</tbody>
</table>

➢ Communality of variable 2 is 45.2%.
   • It implies that the only 45.2% of variation in variable 2 is captured by our extracted factors.
   • This may also partially explain why variable 2 is not appearing in our final interpretation of the table.
Confirmatory Factor Analysis

Confirmatory factor analysis (CFA) starts with a hypothesis about how many factors there are and which items load on which factors. Factor loadings and factor correlations are obtained as in EFA. EFA, in contrast, does not specify a measurement model initially and usually seeks to discover the measurement model. In EFA, all items load on all factors. In CFA, most researchers start with a model in which items load on only one factor (simple structure).
CONFIRMATORY FACTOR ANALYSIS

Confirmatory Factor Analysis

Consistency of Interest

Perseverance of Interest


CONFIRMATORY FACTOR ANALYSIS

Confirmatory Factor Analysis

A test is computed to investigate how well the hypothesized factor structure fits with the data.
The fit test seeks to find a non-significant result, indicating good fit to the data.
CONFIRMATORY FACTOR ANALYSIS

Confirmatory Factor Analysis

The model fit is derived from comparing the correlations (technically, the covariances) among the items to the correlations expected by the model being tested.

Mathematically, certain models imply certain correlations, e.g.,

- If one-factor model, items should be highly correlated, items that do not correlate highly will lead to a poor fit for a one factor model.
- If model specifies that two factors are uncorrelated, then the model will not fit well if items from one factor tend to be correlated with items from another factor.
CONFIRMATORY FACTOR ANALYSIS

Confirmatory Factor Analysis

You may hear about many fit indices, so here are some common examples:

- Chi-square, $\chi^2$, lower values indicate better fit
- RMSEA, lower values indicate better fit ($< .06$)
- SRMR, lower values indicate better fit ($< .08$)
- Comparative Fit Index, higher value indicate better fit ($>.95$)
- Tucker-Lewis Index, higher value indicate better fit ($>.95$)
CONFIRMATORY FACTOR ANALYSIS

Confirmatory Factor Analysis

If the model does not fit well, it can be altered and retested. Many possible loading structures can be specified by the user and any item can load on multiple factors if desired. The more changes made, the more the researcher may be capitalizing on chance, running the risk of Type I errors.
CONFIRMATORY FACTOR ANALYSIS

Confirmatory Factor Analysis

Diagram: Two circles labeled "Consistency of Interest" and "Perseverance of Interest" connected by a line. Three rectangular boxes labeled $X_1$, $X_2$, and $X_3$ are connected to the circles with arrows, labeled "e". Another three boxes labeled $X_1$, $X_2$, and $X_3$ are also connected with arrows labeled "e".
CONFIRMATORY FACTOR ANALYSIS

Confirmatory Factor Analysis

The differences between EFA and CFA are often overstated
Despite their names, both can be used in an exploratory manner
  CFA models can be modified if the model does not fit well
  EFA is sometimes used by researchers even though they have a well-developed idea about the factor structure and wants to confirm it
Both methods are based on discovering number of underlying factors for a set of items and estimating how strongly they relate to the factors
CONFIRMATORY FACTOR ANALYSIS

Confirmatory Factor Analysis

Both methods of factor analysis are sensitive psychometric analysis that provide information about reliability, item quality, and validity.

Scale may be modified by eliminating items or changing the structure of the measure.

Either method may be used as a preliminary step to evaluate a measure or set of subscales that will be computed and used in later research.
CONFIRMATORY FACTOR ANALYSIS

Confirmatory Factor Analysis

Specialized software usually required (e.g., Amos, Mplus, LISREL, EQS, the R package lavaan)

EFA procedures usually available in general statistical software packages like SPSS, SAS, Stata etc.
CONFIRMATORY FACTOR ANALYSIS

Confirmatory Factor Analysis

CFA is part of a larger analysis framework, called *structural equation modeling* (SEM), which combines CFA with path analysis (regression slopes)

SEM can use factors (or “latent variables”) in regression analysis to predict other variables or be predicted by other variables, with the advantage of estimating and eliminating measurement error from correlation and regression estimates
CONFIRMATORY FACTOR ANALYSIS

Confirmatory Factor Analysis

Diagram showing relationships between Positive Social Control, Positive Emotional Responses, Negative Emotional Responses, Intention, and Physical Activity.
CONFIRMATORY FACTOR ANALYSIS

Confirmatory Factor Analysis
CONVERGENT VALIDITY

Validity refers to the truthfulness of a measure
Does it measure what is intends to measure

Degree to which the operationalization is similar to (converges on) other operationalizations to which it theoretically should be similar

Convergent principle

Measures of constructs that are related to each other should be strongly correlated. The correlation provide evidence that the items all converge on the same construct.
DISCRIMINANT VALIDITY (DIVERGENT)

Degree to which the operationalization is not similar to (diverge from) other operationalization to which it theoretically should not be similar

Discriminant Principal

Measures of different constructs should not correlate highly with each other
Note: Measured variables are shown as a box with labels corresponding to those shown in the HBAT questionnaire. Latent constructs are an oval. Each measured variable has an error term, but the error terms are not shown. Two headed connections indicate covariance between constructs. One headed connectors indicate a causal path from a construct to an indicator (measured) variable. In CFA all connectors between constructs are two-headed covariances / correlations.
Factor Loadings – Convergent Validity

These are factor loadings but in AMOS they are called “standardized” regression weights.

Factor loadings are the first thing to look at in examining convergent validity. Our guidelines are that all loadings should be at least .5, and preferably .7 or higher. All loadings are significant as required for convergent validity. The lowest is .592 (OC1) and there are only two below .70 (EP1 & OC3).

When examining convergent validity, we look at two additional measures:

(1) Average Variance Extracted (AVE) by each construct.
(2) Construct Reliabilities (CR).

The AVE and CR are not provided by AMOS software so they have to be calculated.
### CFA Three Factor Completely Standardized Factor Loadings, Variance Extracted, and Reliability Estimates

<table>
<thead>
<tr>
<th>Item</th>
<th>OC Reliabilities</th>
<th>EP Reliabilities</th>
<th>AC Reliabilities</th>
<th>Item Reliabilities</th>
<th>delta</th>
</tr>
</thead>
<tbody>
<tr>
<td>OC1</td>
<td>0.59</td>
<td>0.349</td>
<td>0.65</td>
<td>0.349</td>
<td>0.65</td>
</tr>
<tr>
<td>OC2</td>
<td>0.87</td>
<td>0.759</td>
<td>0.24</td>
<td>0.759</td>
<td>0.24</td>
</tr>
<tr>
<td>OC3</td>
<td>0.67</td>
<td>0.448</td>
<td>0.55</td>
<td>0.448</td>
<td>0.55</td>
</tr>
<tr>
<td>OC4</td>
<td>0.84</td>
<td>0.709</td>
<td>2.264</td>
<td>0.709</td>
<td>2.264</td>
</tr>
<tr>
<td>EP1</td>
<td>0.69</td>
<td>0.477</td>
<td>0.52</td>
<td>0.477</td>
<td>0.52</td>
</tr>
<tr>
<td>EP2</td>
<td>0.81</td>
<td>0.658</td>
<td>0.34</td>
<td>0.658</td>
<td>0.34</td>
</tr>
<tr>
<td>EP3</td>
<td>0.77</td>
<td>0.596</td>
<td>0.40</td>
<td>0.596</td>
<td>0.40</td>
</tr>
<tr>
<td>EP4</td>
<td>0.82</td>
<td>0.679</td>
<td>2.410</td>
<td>0.679</td>
<td>2.410</td>
</tr>
<tr>
<td>AC1</td>
<td>0.82</td>
<td>0.676</td>
<td>0.32</td>
<td>0.676</td>
<td>0.32</td>
</tr>
<tr>
<td>AC2</td>
<td>0.82</td>
<td>0.674</td>
<td>0.33</td>
<td>0.674</td>
<td>0.33</td>
</tr>
<tr>
<td>AC3</td>
<td>0.84</td>
<td>0.699</td>
<td>0.30</td>
<td>0.699</td>
<td>0.30</td>
</tr>
<tr>
<td>AC4</td>
<td>0.82</td>
<td>0.666</td>
<td>0.33</td>
<td>0.666</td>
<td>0.33</td>
</tr>
</tbody>
</table>

#### Variance Extracted

<table>
<thead>
<tr>
<th>Construct</th>
<th>Variance Extracted</th>
</tr>
</thead>
<tbody>
<tr>
<td>OC</td>
<td>56.61%</td>
</tr>
<tr>
<td>EP</td>
<td>60.25%</td>
</tr>
<tr>
<td>AC</td>
<td>67.86%</td>
</tr>
</tbody>
</table>

#### Construct Reliability

<table>
<thead>
<tr>
<th>Construct</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>OC</td>
<td>0.84</td>
</tr>
<tr>
<td>EP</td>
<td>0.86</td>
</tr>
<tr>
<td>AC</td>
<td>0.89</td>
</tr>
</tbody>
</table>

**Factor Loadings**

This is the same as the eigenvalue in exploratory factor analysis.

**Squared Factor Loadings (communalities)**

The delta is calculated as 1 minus the item reliability, e.g., the AC4 delta is $1 - 0.666 = 0.33$

The delta is also referred to as the standardized error variance.
Formula for Variance Extracted

\[ VE = \frac{\sum_{i=1}^{n} \lambda_i^2}{n} \]

In the formula above the \( \lambda \) represents the standardized factor loading and \( i \) is the number of items. So, for \( n \) items, AVE is computed as the sum of the squared standardized factor loadings divided by the number of items, as shown above.

A good rule of thumb is an AVE of .5 or higher indicates adequate convergent validity. An AVE of less than .5 indicates that on average, there is more error remaining in the items than there is variance explained by the latent factor structure you have imposed on the measure.

An AVE estimate should be computed for each latent construct in a measurement model.

Calculated Variance Extracted (AVE):

- OC Construct = .349 + .759 + .448 + .709 = 2.264 / 4 = .5661
- AC Construct = .676 + .674 + .699 + .666 = 2.714 / 4 = .6786
Formula for Construct Reliability

\[ CR = \frac{\left( \sum_{i=1}^{n} \lambda_i \right)^2}{\left( \sum_{i=1}^{n} \lambda_i \right)^2 + \left( \sum_{i=1}^{n} \delta_i \right)} \]

The sum of the loadings, squared

Computation of Construct Reliability (CR)

The sum of the loadings, squared

The sum of the error variance (delta)

Construct reliability — is computed from the sum of factor loadings (\( \lambda_i \)), squared for each construct and the sum of the error variance terms for a construct (\( \delta_i \)) using the above formula. Note: error variance is also referred to as delta.

The rule of thumb for a construct reliability estimate is that .7 or higher suggests good reliability. Reliability between .6 and .7 may be acceptable provided that other indicators of a model’s construct validity are good. A high construct reliability indicates that internal consistency exists. This means the measures all are consistently representing something.
Evaluation Three-Construct Model
Convergent Validity

Taken together, the evidence provides initial support for the convergent validity of the three construct HBAT measurement model. Although three loading estimates are below .7, two of these are just below the .7 and do not appear to be significantly harming model fit or internal consistency.

The average variance extracted (AVE) estimates all exceed .5 and the construct reliability estimates all exceed .7. In addition, the model fits relatively well based on the GOF measures. Therefore, all the indicator items are retained at this point and adequate evidence of convergent validity is provided.

We now move on to examine:

(1) Discriminant validity
(2) Nomological validity
Discriminant Validity

Discriminant validity = the extent to which a construct is truly distinct from other constructs.

Rule of Thumb: all construct average variance extracted (AVE) estimates should be larger than the corresponding squared interconstruct correlation estimates (SIC). If they are, this indicates the measured variables have more in common with the construct they are associated with than they do with the other constructs.
Correlations between the EP, AC and OC constructs. These are standardized covariances. These are used in calculating discriminant validity.

Covariances between the EP, AC and OC constructs.

Correlations between the EP, AC and OC constructs. These are standardized covariances. These are used in calculating discriminant validity.
Discriminant Validity

In the columns below, we calculate the SIC (Squared Interconstruct Correlations) from the IC (Innerconstruct Correlations) obtained from the correlations table on the AMOS printout (see previous slide):

<table>
<thead>
<tr>
<th>IC</th>
<th>SIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>EP – AC</td>
<td>.254</td>
</tr>
<tr>
<td>EP – OC</td>
<td>.500</td>
</tr>
<tr>
<td>AC – OC</td>
<td>.303</td>
</tr>
</tbody>
</table>

Discriminant validity - compares the average variance extracted (AVE) estimates for each factor with the squared interconstruct correlations (SIC) associated with that factor, as shown below:

<table>
<thead>
<tr>
<th></th>
<th>AVE</th>
<th>SIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>OC Construct</td>
<td>.5661</td>
<td>.2500, .0918</td>
</tr>
<tr>
<td>EP Construct</td>
<td>.6025</td>
<td>.0645, .2500</td>
</tr>
<tr>
<td>AC Construct</td>
<td>.6786</td>
<td>.0645, .0918</td>
</tr>
</tbody>
</table>

All variance extracted (AVE) estimates in the above table are larger than the corresponding squared interconstruct correlation estimates (SIC). This means the indicators have more in common with the construct they are associated with than they do with other constructs. Therefore, the HBAT three construct CFA model demonstrates discriminant validity.
THANK YOU